Guarantees of Total Variation Minimization for Signal Recovery

Jian-Feng Cai and Weiyu Xu

March 21, 2013

Abstract

In this paper, we consider using total variation minimization to recover signals whose gradients have a sparse support, from a small number of measurements. We establish the proof for the performance guarantee of total variation (TV) minimization in recovering *one-dimensional* signal with sparse gradient support. This partially answers the open problem of proving the fidelity of total variation minimization in such a setting [20]. In particular, we have shown that the recoverable gradient sparsity can grow linearly with the signal dimension when TV minimization is used. Recoverable sparsity thresholds of TV minimization are explicitly computed for 1-dimensional signal by using the Grassmann angle framework. We also extend our results to TV minimization for multidimensional signals.

1 Introduction

Compressed sensing has recently gained a lot of attention in many applications including medical imaging, because it enables acquiring sparse signals from a much smaller number of samples than the ambient dimension of signal. Compressed sensing takes advantage of the fact that most signals of interest in practice are sparse: there are only a few nonzero or big elements when the signals are represented over a certain dictionary such as wavelet basis. For these types of sparse or compressible signals, compressed sensing theory [5, 10] has established that a small number of nonadaptive measurements are often sufficient to efficiently recover them under methods such as ℓ_1 minimization [4–6, 10].

Without of loss of generality, let us assume that $\boldsymbol{x} \in \mathbb{R}^N$ is a one-dimensional (compared with 2-dimensional images and 3-dimensional videos) signal vector of N elements, and has no more than K ($K \ll N$) nonzero elements. In compressed sensing, we sample $\boldsymbol{x}$ using M ($M < N$) linear projections

$$\boldsymbol{y} = A\boldsymbol{x},$$

where A is an $M \times N$ measurement matrix and $\boldsymbol{y}$ is an $M \times 1$ measurement result vector. Knowing A and the measurement result $\boldsymbol{y}$, ℓ_1 minimization is often used to recover the sparse $\boldsymbol{x}$:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \qquad \text{subject to} \;\; \boldsymbol{y} = A\boldsymbol{x}. \tag{1}$$

It has been shown that under suitable conditions on the measurement matrix A, it is guaranteed that the original $\boldsymbol{x}$ is the unique solution to ℓ_1 minimization (1). In fact, if A satisfies the so-called restricted isometry property (RIP), then the solution of (1) matches exactly with the original signal [2, 5, 13]. Various results concerning the perfect reconstruction of the original signal by solving (1) have been established in [5, 8, 10–14, 17, 28, 30].

The results above hold true only for sparse signals, and they can be extended to signals that are synthesized by a linear combination of few atoms in a (redundant) dictionary with incoherent atoms [21]. However, there are numerous practical examples in which a signal of interest does not fall into the category where the aforementioned theory work. One such an example is signal that has a sparse gradient (i.e., the signal is piecewise constant), which arises frequently from imaging. Images with little detail are usually modelled as piecewise constant functions. For simplicity, we assume that $\boldsymbol{x} \in \mathbb{R}^N$ is a vector generated from 1-dimensional

piecewise constant signal. Let $D\boldsymbol{x}$ be its finite difference defined by $[D\boldsymbol{x}]_i = x_{i+1} - x_i$ for $i = 1, 2, \ldots, N-1$. Since $\boldsymbol{x}$ is piecewise constant, we must have that $D\boldsymbol{x}$ is sparse. Assume that $D\boldsymbol{x}$ has only K ($K \ll N$) nonzero entries. Let $\boldsymbol{y} = A\boldsymbol{x} \in \mathbb{R}^M$ be M linear samples of $\boldsymbol{x}$. Then, to recover $\boldsymbol{x}$, one usually solves

$$\min_{\boldsymbol{x}} \|D\boldsymbol{x}\|_1 \qquad \text{subject to} \quad \boldsymbol{y} = A\boldsymbol{x}. \tag{2}$$

The regularization term $\|D\boldsymbol{x}\|_1$ is called the *total variation (TV)* of $\boldsymbol{x}$. When $\boldsymbol{x} \in \mathbb{R}^{N^d}$ is generated from d-dimensional signals, we only need to replace D by the concatenation of directional finite differences, and $\|D\boldsymbol{x}\|_1$ is the anisotropic TV of $\boldsymbol{x}$.

TV regularization has been used extensively in the literature for decades in imaging sciences [1,18,23,24] and other related fields [7,26]. The minimization problem (2) has the same form as the minimization in the analysis-based compressed sensing in [3]. However, the perfect reconstruction result in [3] can not be applied to (2), as the rows of D do not form a frame (D has a nontrivial null space). Despite the great importance of the TV minimization in applications, rigorous proofs of conditions of successfully recovering signal by using the TV minimization have only recently been established [19,20]. To establish such conditions, [19,20] first transformed d-dimensional ($d \geq 2$) signals with sparse gradients into signals compressible over the Haar orthogonal wavelet basis. Then a modified restricted isometry condition, which takes into account the Haar orthogonal wavelet transformation, was established for the matrix A such that (2) offers a stable recovery of $\boldsymbol{x}$. However, it is noted in [19, 20] that establishing conditions for successfully recovering 1-dimensional (namely d=1) signal vector remains an open problem. This is partially due to the fact that small TV of a 1-dimensional signal does not necessarily imply fast decay of its Haar wavelet coefficients.

In this paper, we establish the proof for performance guarantees of TV minimization in recovering 1-*dimensional* signal with sparse gradient support. This partially answers the open problem of proving the fidelity of total variation minimization in such a setting [20]. Compared with [19,20], our results do not use the restricted isometry condition, but instead directly work on the null space condition of the measurement matrix A. To establish the null space condition of interest, we use "Escape through the Mesh" theorem [16,22,25] to estimate the Gaussian width [16,22] of a cone specified by the null space condition. We then use the Grassmann angle framework to calculate the thresholds on gradient sparsity such that TV minimization (2) can recover with high probability. We further extend our results to TV minimization for higher dimensional signals. For $d \geq 2$, we have obtained performance bounds for TV minimization comparable to results in [19,20]. In [9], an average-case phase transition was calculated for TV minimization through evaluating the asymptotic minimax Mean Square Error (MSE) of TV minimization. Compared with [9], our results are more concerned with the worst-case performance guarantees which are uniformly true for all the possible supports for the signal gradient.

The rest of this paper is organized as follows. In Section 2, we establish the performance guarantee of TV minimization for 1-dimensional signal vector. Our proof is based on a null space condition introduced in Section 2.1, and two different arguments are respectively presented in Section 2.2 using the escape through the mesh theorem and in Section 2.3 via Grassman angle framework. In Section 3, we extend our results to TV minimization for multidimensional signals. Section 4 concludes our paper and discusses future directions.

2 One-dimensional Signals

In this section, we establish the main result of this paper on the performance guarantee of TV minimization in recovering one-dimensional signal with sparse gradient support. Throughout this section, we will assume that $\boldsymbol{x} \in \mathbb{R}^N$ is generated from a one-dimensional signal and $D\boldsymbol{x}$ contains at most K nonzeros. We assume that the entries of $A \in \mathbb{R}^{M \times N}$ are randomly drawn from i.i.d. Gaussian distribution. We give two different arguments on the proofs, namely, the one using "Escape through the Mesh" theorem [16,22,25] in Section 2.2 and the Grassmann angle framework in Section 2.3. There two arguments will lead to two different recovery threshold bounds for minimal M. Both the arguments are based on a null space property of the matrix A, which are presented in Section 2.1.

2.1 The Null Space Condition for Successful Recovery via the TV Minimization

In this section, we give a condition on the null space of the linear projection matrix A, such that TV minimization successfully recovers one-dimensional signals with sparse gradients. We remark that this condition is not new, and it has appeared in the proofs in [19, 20].

Theorem 2.1. *Assume $A \in \mathbb{R}^{M\times N}$ and $\boldsymbol{y} = A\boldsymbol{x}$. Then $\boldsymbol{x}$ is the unique solution to (2) for all $\boldsymbol{x}$ whose gradients $D\boldsymbol{x}$ have no more than K nonzero elements (no matter what the support $\mathcal{K}$ of $D\boldsymbol{x}$ is) if and only if the following condition holds: for every nonzero vector $\boldsymbol{z}$ in the null space of A (namely $A\boldsymbol{z} = 0$, $\boldsymbol{z} \neq \boldsymbol{0}$),*

$$\|(D\boldsymbol{z})_{\mathcal{K}}\|_1 < \|(D\boldsymbol{z})_{\mathcal{K}^c}\|_1 \quad \forall \mathcal{K} \subset \{1, 2, \dots, N-1\} \ s.t.\ |\mathcal{K}| \le K. \tag{3}$$

We omit the detailed proof of this theorem, since it is very similar to the proof of null space conditions for ℓ_1 minimization; see, for example, [25, 30].

2.2 Recovery Thresholds via Escape through the Mesh Theorem

In this subsection, we prove that a measurement matrix A whose elements are i.i.d. Gaussian random variables satisfies the null space condition in Theorem 2.1 with high probability, as long as

$$M \ge C(NK)^{1/2} \ln N,$$

where $C > 0$ is a constant. This shows that, for any M that is proportionally growing with N, the signal gradient sparsity K that the TV minimization is guaranteed to recover will also grow proportionally with N. Our proof builds on the following "Escape through the Mesh" theorem.

Theorem 2.2 (Escape through the mesh [16])**.** *Let $\mathcal{S}$ be a subset of the unit Euclidean sphere $\mathbb{S}^{N-1}$ in $\mathbb{R}^N$. Let Y be a random $(N-M)$-dimensional subspace of $\mathbb{R}^N$, distributed uniformly in the Grassmanian with respect to the Haar measure. Define the Gaussian width for the set $\mathcal{S}$ as $w(\mathcal{S})=E(\sup_{\boldsymbol{w}\in\mathcal{S}}(\boldsymbol{h}^T\boldsymbol{w}))$, where $\boldsymbol{h}$ is a random column vector in $\mathbb{R}^N$ with i.i.d. $\mathcal{N}(0,1)$ Gaussian elements. Assume that $w(\mathcal{S}) < (\sqrt{M} - \frac{1}{2\sqrt{M}})$. Then*

$$P(Y \bigcap S = \emptyset) > 1 - 3.5e^{-\frac{(\sqrt{M}-\frac{1}{2\sqrt{M}})-w(\mathcal{S})}{18}}.$$

If the elements of the measurement matrix A are i.i.d. Gaussian random variables, then the null space of A is a random $(N-M)$-dimensional subspace distributed uniformly in the Grassmanian with respect to the Haar measure (see [25]). To prove the null space condition in Theorem 2.1 holds with high probability, we show that the Gaussian width $w(\mathcal{S})$ is in the order of $\sqrt{M}$ for the set

$$\mathcal{S} = \{\boldsymbol{x} \ : \ \|\boldsymbol{x}\|_2 = 1, \quad \text{and} \quad \exists \mathcal{K} \subset \{1, 2, \dots, N\} \text{ s.t. } |\mathcal{K}| \le K,\ \|(D\boldsymbol{x})_{\mathcal{K}}\|_1 \ge \|(D\boldsymbol{x})_{\mathcal{K}^c}\|_1\}.$$

For any $\boldsymbol{x} \in \mathcal{S}$ and a set $\mathcal{K}$ that satisfy $\|(D\boldsymbol{x})_{\mathcal{K}}\|_1 \ge \|(D\boldsymbol{x})_{\mathcal{K}^c}\|_1$, we have that

$$\|(D\boldsymbol{x})_{\mathcal{K}^c}\|_1 \le \|(D\boldsymbol{x})_{\mathcal{K}}\|_1 \le \sqrt{K}\|(D\boldsymbol{x})_{\mathcal{K}}\|_2 \le \sqrt{K}\|D\boldsymbol{x}\|_2 \le 2\sqrt{K}\|\boldsymbol{x}\|_2 = 2\sqrt{K}.$$

This implies $\|D\boldsymbol{x}\|_1 \le 4\sqrt{K}$ and therefore and further

$$\mathcal{S} \subset \widetilde{\mathcal{S}} := \{\boldsymbol{x} \ : \ \|\boldsymbol{x}\|_2 \le 1, \ \ \|D\boldsymbol{x}\|_1 \le 4\sqrt{K}\}.$$

In the following, we estimate the Gaussian width of $\widetilde{\mathcal{S}}$. We only consider the case that $N = 2^L$, and the proof the other cases are essentially the same and does not change only the order the the Gaussian width.

For any $\boldsymbol{x} \in \widetilde{\mathcal{S}}$, we decompose $\boldsymbol{x}$ according to Haar wavelet transform as

$$\boldsymbol{x} = \hat{\boldsymbol{z}}^{(1)} + \ldots + \hat{\boldsymbol{z}}^{(L)} + \hat{\boldsymbol{y}}^{(L)}, \tag{4}$$

where

$$\hat{\boldsymbol{z}}^{(\ell)} = \boldsymbol{z}^{(\ell)} \otimes [\underbrace{1 \ \ldots \ 1}_{2^{\ell-1}} \ \underbrace{-1 \ \ldots \ -1}_{2^{\ell-1}}], \quad \boldsymbol{z}^{(\ell)} = [z_1^{(\ell)} \ z_2^{(\ell)} \ \ldots \ z_{N/2^\ell}^{(\ell)}]$$

and

$$\hat{\boldsymbol{y}}^{(L)} = \boldsymbol{y}^{(L)} \otimes [1 \ 1 \ \ldots \ 1], \quad \boldsymbol{y}^{(L)} = [y_1^{(L)}].$$

Here $\otimes$ is the Kronecker product, i.e., $\boldsymbol{a} \otimes \boldsymbol{b} := [a_1\boldsymbol{b} \ a_2\boldsymbol{b} \ \ldots \ a_n\boldsymbol{b}]$. The decomposition (4) is done recursively as follows. We first decompose $\boldsymbol{x} = \hat{\boldsymbol{y}}^{(1)} + \hat{\boldsymbol{z}}^{(1)}$, where

$$\hat{\boldsymbol{y}}^{(1)} = \boldsymbol{y}^{(1)} \otimes [1 \ 1], \qquad \boldsymbol{y}^{(1)} = [y_1^{(1)} \ y_2^{(1)} \ \ldots \ y_{N/2}^{(1)}], \quad y_i^{(1)} = \frac{x_{2i-1} + x_{2i}}{2},$$

and

$$\hat{\boldsymbol{z}}^{(1)} = \boldsymbol{z}^{(1)} \otimes [1 \ -1], \qquad \boldsymbol{z}^{(1)} = [z_1^{(1)} \ z_2^{(1)} \ \ldots \ z_{N/2}^{(1)}], \quad z_i^{(1)} = \frac{x_{2i-1} - x_{2i}}{2}.$$

Then, we further decompose

$$\hat{\boldsymbol{y}}^{(1)} = \hat{\boldsymbol{y}}^{(2)} + \hat{\boldsymbol{z}}^{(2)},$$

where

$$\hat{\boldsymbol{y}}^{(2)} = \boldsymbol{y}^{(2)} \otimes [1 \ 1 \ 1 \ 1], \qquad \boldsymbol{y}^{(2)} = [y_1^{(2)} \ y_2^{(2)} \ \ldots \ y_{N/4}^{(2)}], \quad y_i^{(2)} = \frac{y_{2i-1}^{(1)} + y_{2i}^{(1)}}{2},$$

and

$$\hat{\boldsymbol{z}}^{(2)} = \boldsymbol{z}^{(2)} \otimes [1 \ 1 \ -1 \ -1], \qquad \boldsymbol{z}^{(2)} = [z_1^{(2)} \ z_2^{(2)} \ \ldots \ z_{N/4}^{(2)}], \quad z_i^{(2)} = \frac{y_{2i-1}^{(1)} - y_{2i}^{(1)}}{2}.$$

Generally, at level ℓ, we have that

$$\hat{\boldsymbol{y}}^{(\ell)} = \boldsymbol{y}^{(\ell)} \otimes [\underbrace{1 \ \ldots \ 1}_{2^\ell}], \qquad \boldsymbol{y}^{(\ell)} = [y_1^{(\ell)} \ y_2^{(\ell)} \ \ldots \ y_{N/2^\ell}^{(\ell)}]$$

we decompose it as

$$\hat{\boldsymbol{y}}^{(\ell)} = \hat{\boldsymbol{y}}^{(\ell+1)} + \hat{\boldsymbol{z}}^{(\ell+1)},$$

where

$$\hat{\boldsymbol{y}}^{(\ell+1)} = \boldsymbol{y}^{(\ell+1)} \otimes [\underbrace{1 \ \ldots \ 1}_{2^{\ell+1}}], \qquad \boldsymbol{y}^{(\ell+1)} = [y_1^{(\ell+1)} \ z_2^{(\ell+1)} \ \ldots \ y_{N/2^{\ell+1}}^{(\ell+1)}], \quad y_i^{(\ell+1)} = \frac{y_{2i-1}^{(\ell)} + y_{2i}^{(\ell)}}{2},$$

and

$$\hat{\boldsymbol{z}}^{(\ell+1)} = \boldsymbol{z}^{(\ell+1)} \otimes [\underbrace{1 \ \ldots \ 1}_{2^\ell} \ \underbrace{-1 \ \ldots \ -1}_{2^\ell}], \qquad \boldsymbol{z}^{(\ell+1)} = [z_1^{(\ell+1)} \ z_2^{(\ell+1)} \ \ldots \ z_{N/2^{\ell+1}}^{(\ell+1)}], \quad z_i^{(\ell+1)} = \frac{y_{2i-1}^{(\ell)} - y_{2i}^{(\ell)}}{2}.$$

The decomposition (4) has the following properties.

- Obviously, components in decomposition (4) are orthogonal to each others. Consequently,

$$\|\boldsymbol{x}\|_2^2 = \|\hat{\boldsymbol{z}}^{(1)}\|_2^2 + \|\hat{\boldsymbol{z}}^{(2)}\|_2^2 + \ldots + \|\hat{\boldsymbol{z}}^{(L)}\|_2^2 + \|\hat{\boldsymbol{y}}^{(L)}\|_2^2 = \sum_{\ell=1}^{L} \left(2^\ell \|\boldsymbol{z}^{(\ell)}\|_2^2\right) + 2^L \|\boldsymbol{y}\|_2^2.$$

  Since $\boldsymbol{x} \in \tilde{\mathcal{S}}$ implies $\|\boldsymbol{x}\|_2^2 \leq 1$, we have

$$\sum_{\ell=1}^{L} \left(2^\ell \|\boldsymbol{z}^{(\ell)}\|_2^2\right) + 2^L \|\boldsymbol{y}^{(L)}\|_2^2 \leq 1. \tag{5}$$

- It can be shown that
$$\|D\hat{\boldsymbol{y}}^{(\ell)}\|_1 \le \|D\hat{\boldsymbol{y}}^{(\ell-1)}\|_1$$
and, therefore,
$$\|\boldsymbol{z}^{(\ell)}\|_1 \le \|D\hat{\boldsymbol{y}}^{(\ell-1)}\|_1/2 \le 2\sqrt{K}. \tag{6}$$
Indeed, let $\boldsymbol{u}$ be satisfying $\|D\hat{\boldsymbol{y}}^{(\ell)}\|_1 = \langle \boldsymbol{u}, D\hat{\boldsymbol{y}}^{(\ell)}\rangle$ and $\|\boldsymbol{u}\|_\infty \le 1$, and we then have
$$\begin{aligned}
\langle \boldsymbol{u}, D\hat{\boldsymbol{y}}^{(\ell)}\rangle &= \sum_{i=1}^{N/2^\ell-1}\left(u_{i2^\ell}\cdot(y_{i+1}^{(\ell)}-y_i^{(\ell)})\right) = \sum_{i=1}^{N/2^\ell-1}\left((u_{i2^\ell})\cdot\left(\frac{y_{2i+2}^{(\ell-1)}+y_{2i+1}^{(\ell-1)}}{2}-\frac{y_{2i}^{(\ell-1)}+y_{2i-1}^{(\ell-1)}}{2}\right)\right)\\
&= \sum_{i=1}^{N/2^\ell-1}\left(u_{i2^\ell}\cdot\left(\frac{y_{2i+2}^{(\ell-1)}-y_{2i+1}^{(\ell-1)}}{2}+(y_{2i+1}^{(\ell-1)}-y_{2i}^{(\ell-1)})+\frac{y_{2i}^{(\ell-1)}-y_{2i-1}^{(\ell-1)}}{2}\right)\right)\\
&= \sum_{i=1}^{N/2^\ell-1}\left(u_{i2^\ell}\cdot\left(\frac{y_{2i+2}^{(\ell-1)}-y_{2i+1}^{(\ell-1)}}{2}\right)\right)+\sum_{i=1}^{N/2^\ell-1}\left(u_{i2^\ell}\cdot(y_{2i+1}^{(\ell-1)}-y_{2i}^{(\ell-1)})\right)\\
&\quad+\sum_{i=1}^{N/2^\ell-1}\left(u_{i2^\ell}\cdot\left(\frac{y_{2i}^{(\ell-1)}-y_{2i-1}^{(\ell-1)}}{2}\right)\right)\\
&= \frac{u_{2^\ell}}{2}\cdot\left(y_2^{(\ell-1)}-y_1^{(\ell-1)}\right)+\sum_{i=2}^{N/2^\ell-1}\left(\frac{u_{i2^\ell}+u_{(i-1)2^\ell}}{2}\cdot\left(y_{2i}^{(\ell-1)}-y_{2i-1}^{(\ell-1)}\right)\right)\\
&\quad+\frac{u_{N-2^\ell}}{2}\cdot\left(y_{N/2^{\ell-1}}^{(\ell-1)}-y_{N/2^{\ell-1}-1}^{(\ell-1)}\right)+\sum_{i=1}^{N/2^\ell-1}\left(u_{i2^\ell}\cdot(y_{2i+2}^{(\ell-1)}-y_{2i+1}^{(\ell-1)})\right)\\
&= \langle\tilde{\boldsymbol{u}}, D\hat{\boldsymbol{y}}^{(\ell-1)}\rangle,
\end{aligned}$$
where
$$\tilde{\boldsymbol{u}} = \left[\boldsymbol{0}\ \frac{u_{2^\ell}}{2}\ \boldsymbol{0}\ u_{2^\ell}\ \boldsymbol{0}\ \frac{u_{22^\ell}+u_{2^\ell}}{2}\ \boldsymbol{0}\ u_{22^\ell}\ \boldsymbol{0}\ \frac{u_{32^\ell}+u_{22^\ell}}{2}\ \boldsymbol{0}\ u_{32^\ell}\ \boldsymbol{0}\ \ldots\ldots\ldots\ \boldsymbol{0}\ \frac{u_{N-2^\ell}}{2}\ \boldsymbol{0}\right].$$
Since
$$\left|\frac{u_{i2^\ell}+u_{(i-1)2^\ell}}{2}\right| \le \frac{|u_{i2^\ell}|+|u_{(i-1)2^\ell}|}{2},$$
we have $\|\tilde{\boldsymbol{u}}\|_\infty \le \|\boldsymbol{u}\|_\infty \le 1$. This leads to
$$\|D\hat{\boldsymbol{y}}^{(\ell)}\|_1 = \langle\boldsymbol{u}, D\hat{\boldsymbol{y}}^{(\ell)}\rangle = \langle\tilde{\boldsymbol{u}}, D\hat{\boldsymbol{y}}^{(\ell-1)}\rangle \le \|D\hat{\boldsymbol{y}}^{(\ell-1)}\|_1.$$

Now we are ready to estimate the Gaussian width of $\tilde{\mathcal{S}}$. Let $\boldsymbol{g}$ be a vector whose entries are i.i.d. Gaussian random variables with mean 0 and variance 1. Since (5) implies $\|\boldsymbol{z}^{(\ell)}\|_2 \le \frac{1}{\sqrt{2^\ell}}$, we have, by Cauchy-Schwartz inequality, $\|\boldsymbol{z}^{(\ell)}\|_1 \le \sqrt{\frac{N}{2^\ell}}\|\boldsymbol{z}^{(\ell)}\|_2 \le \frac{\sqrt{N}}{2^\ell}$. This together with (6) implies that
$$\|\boldsymbol{z}^{(\ell)}\|_1 \le \min\left\{\frac{\sqrt{N}}{2^\ell}, 2\sqrt{K}\right\}.$$
Then,
$$\langle\hat{\boldsymbol{z}}^{(\ell)}, \boldsymbol{g}\rangle = \langle\boldsymbol{z}^{(\ell)}, \boldsymbol{g}^{(\ell)}\rangle \le \|\boldsymbol{z}^{(\ell)}\|_1\|\boldsymbol{g}^{(\ell)}\|_\infty.$$
Here
$$\boldsymbol{g}^{(\ell)} = \left[\sum_{i=1}^{2^{\ell-1}}(g_i - g_{i+2^{\ell-1}})\ \sum_{i=1}^{2^{\ell-1}}(g_{i+2^\ell}-g_{i+2^\ell+2^{\ell-1}})\ \ldots\ \sum_{i=1}^{2^{\ell-1}}(g_{i+N-2^\ell}-g_{i+N-2^\ell+2^{\ell-1}})\right] := [g_1^{(\ell)}\ g_2^{(\ell)}\ \ldots\ g_{N/2^\ell}^{(\ell)}]$$

In the following, we estimate $E(\|\boldsymbol{g}^{(\ell)}\|_\infty)$. Notice that the components in $\boldsymbol{g}^{(\ell)}$ are i.i.d. random variables that follow $\mathcal{N}(0,2^\ell)$. The following argument follows from Lemma 4.4 of Rudelson and Vershynin's paper [22]. Let p be a large enough number that is determined later. Then

$$\begin{aligned} E\left(\|\boldsymbol{g}^{(\ell)}\|_\infty\right) &\le E\left(\left(\sum_i |g_i^{(\ell)}|^p\right)^{1/p}\right) \le \left(N/2^\ell\right)^{1/p}\left(E\left(|g_i^{(\ell)}|^p\right)\right)^{1/p} \\ &\le \sqrt{2^\ell}\left(N/2^\ell\right)^{1/p}\left(2^{p/2}\frac{\Gamma(p/2+1/2)}{\Gamma(1/2)}\right)^{1/p} \\ &\le \sqrt{2^\ell}\left(N/2^\ell\right)^{1/p}\left(\frac{p+1}{e}\right)^{1/2}(1+o(1)). \end{aligned}$$

Choose $p = 2\ln\left(\frac{N/2^\ell}{K}\right)$, and we obtain

$$E\left(\|\boldsymbol{g}^{(\ell)}\|_\infty\right) \le \sqrt{2^\ell}\,(p+1)^{1/2} = \sqrt{2^\ell}\sqrt{1+2\ln\left(N/2^\ell\right)} = \sqrt{2^\ell}\sqrt{2\ln\left(e^{1/2}N/2^\ell\right)}$$

Therefore,

$$\begin{aligned} E\left(\sup_{\boldsymbol{x}\in\tilde{\mathcal{S}}}\langle\hat{\boldsymbol{z}}^\ell,\boldsymbol{g}\rangle\right) &\le E\left(\sup_{\boldsymbol{x}\in\tilde{\mathcal{S}}}\|\hat{\boldsymbol{z}}^\ell\|_1\|\boldsymbol{g}^{(\ell)}\|_\infty\right) \le \min\left\{\frac{\sqrt{N}}{2^\ell}, 2\sqrt{K}\right\}\cdot E\left(\|\boldsymbol{g}^{(\ell)}\|_{(K)}\right) \\ &\le \min\left\{\sqrt{\frac{N}{2^\ell}}, 2\sqrt{2^\ell K}\right\}\cdot\sqrt{2\ln\left(e^{1/2}N/2^\ell\right)}. \end{aligned} \tag{7}$$

This together with (4) implies that

$$E_{\boldsymbol{g}}\left(\sup_{\boldsymbol{x}\in\tilde{\mathcal{S}}}\langle\boldsymbol{x},\boldsymbol{g}\rangle\right) \le \sqrt{\frac{2}{\pi}} + \sum_{\ell=1}^{L}\min\left\{\sqrt{\frac{N}{2^\ell}}, 2\sqrt{2^\ell K}\right\}\cdot\sqrt{2\ln\left(e^{1/2}N/2^\ell\right)}$$

Now we estimate the constant $\sum_{\ell=1}^{L}\min\left\{\sqrt{\frac{N}{2^\ell}}, 2\sqrt{2^\ell K}\right\}\cdot\sqrt{2\ln\left(e^{1/2}N/2^\ell\right)}$. Let L_0 be the maximum integer that satisfies $\sqrt{\frac{N}{2^{L_0}}} \ge 2\sqrt{2^{L_0}K}$, which leads to $2^{L_0} \le \frac{1}{2}\sqrt{\frac{N}{K}}$. Since L_0 is the maximum integer, we have $\frac{1}{2}\sqrt{\frac{N}{K}} \le 2^{L_0+1}$. It is obviously that $\min\left\{\sqrt{\frac{N}{2^\ell}}, 2\sqrt{2^\ell K}\right\} = 2\sqrt{2^\ell K}$ if $\ell \le L_0$ and $\min\left\{\sqrt{\frac{N}{2^\ell}}, 2\sqrt{2^\ell K}\right\} = \sqrt{\frac{N}{2^\ell}}$

otherwise. Therefore, if $N > 1$ and $K > 1$, then

$$
\begin{aligned}
&\sum_{\ell=1}^{L}\left(\min\left\{\sqrt{\frac{N}{2^{\ell}}}, 2\sqrt{2^{\ell}K}\right\}\cdot\sqrt{2\ln\left(e^{1/2}N/2^{\ell}\right)}\right)\\
\leq&\sqrt{2\ln\left(e^{1/2}N\right)}\left(2\sqrt{K}\sum_{\ell=1}^{L_0}\sqrt{2}^{\ell}+\sqrt{N}\sum_{\ell=L_0+1}^{L}\left(\frac{1}{\sqrt{2}}\right)^{\ell}\right)\\
=&\sqrt{2\ln\left(e^{1/2}N\right)}\left(2\sqrt{2}\sqrt{K}\frac{\sqrt{2}^{L_0}-1}{\sqrt{2}-1}+\sqrt{N}\left(\frac{1}{\sqrt{2}}\right)^{L_0+1}\frac{1-\left(\frac{1}{\sqrt{2}}\right)^{L-L_0}}{1-\frac{1}{\sqrt{2}}}\right)\\
\leq&\sqrt{2\ln\left(e^{1/2}N\right)}\left(\left(\frac{2\sqrt{2}}{\sqrt{2}-1}\right)\sqrt{K}\sqrt{2}^{L_0}+\sqrt{\frac{N}{2^{L_0+1}}}\frac{1}{1-\frac{1}{\sqrt{2}}}\right)-\sqrt{\frac{2}{\pi}}\\
\leq&\sqrt{2\ln\left(e^{1/2}N\right)}\left((4+2\sqrt{2})\sqrt{K}\sqrt{\frac{1}{2}\sqrt{\frac{N}{K}}}+(2+\sqrt{2})\sqrt{\frac{N}{\frac{1}{2}\sqrt{\frac{N}{K}}}}\right)-\sqrt{\frac{2}{\pi}}\\
\leq&\left(4\sqrt{2}+4\right)(NK)^{1/4}\sqrt{2\ln\left(e^{1/2}N\right)}-\sqrt{\frac{2}{\pi}}
\end{aligned}
\tag{8}
$$

Finally, we get

$$
E\left(\sup_{\boldsymbol{x}\in\tilde{\mathcal{S}}}\langle\boldsymbol{x},\boldsymbol{g}\rangle\right)=(4\sqrt{2}+4)(NK)^{1/4}\sqrt{2\ln\left(e^{1/2}N\right)}.
$$

Since the Gaussian width in Theorem 2.2 is of the order $\sqrt{M}$, we have

$$
M\sim(NK)^{1/2}\ln N.
$$

2.3 Recovery Thresholds via the Grassmann Angle Framework

In previous subsections, we have used the "Escape through the Mesh" theorem to establish performance guarantees of TV minimization for signal recovery. In this subsection, we explore the Grassmann angle framework [28] to characterize performance guarantees of TV minimization for 1-dimensional signal vectors. The upshot here is that the Grassmann angle framework gives explicitly computable thresholds on recoverable sparsity level K, when the number of measurements is proportionally growing with the signal dimension N.

Let us use $\mathcal{K}$ to denote the set of indices i's such that $|x_{i+1}-x_i|$ is one of the K terms on the left side of the inequality 2.1. Let us denote the set of indices (i's and $(i+1)$'s) involved in these K terms as $\mathcal{DK}$. We note that the cardinality $|\mathcal{DK}|$ of $\mathcal{DK}$ is at most $2K$.

Then there exist at least $(N-1-3K)$ terms in the form of $|x_{i+1}-x_i|$ that do not involve any index in $\mathcal{DK}$. Among these $(N-1-3K)$ terms, we can at least choose $\frac{N-1-3K}{2}$ terms such that each of them involves different indices from $\mathcal{DK}$, *and* from each other. Let us use $\mathcal{KB}$ to denote the set of indices i's such that $|x_{i+1}-x_i|$ is one of these $\frac{N-1-3K}{2}$ terms. By the triangle inequality,

$$
\sum_{i\in\mathcal{K}}|x_{i+1}-x_i|\leq 2\sum_{i\in\mathcal{DK}}|x_i|
$$

Then one sufficient condition for TV minimization to work is

$$
2\sum_{i\in\mathcal{DK}}|x_i|\leq\sum_{i\in\mathcal{KB}}|x_{i+1}-x_i|
\tag{9}
$$

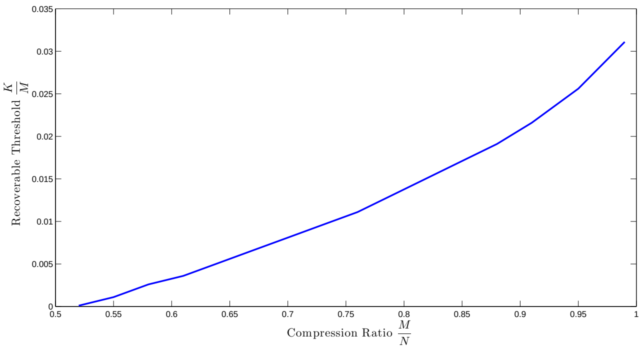

Figure 1: Recoverable thresholds on sparsity of gradient support for TV minimization from the Grassmann angle framework [28]

holds for every vector $\boldsymbol{x}$ in the null space of the projection A. We call this condition $RelaxedNULL$ condition.

Since we are taking the projection A uniformly over all the m-dimensional subspaces in R^n, the probability that $RelaxedNULL$ condition holds, is equivalent to the probability that

$$2\sum_{i\in\{1,2,\dots,|\mathcal{DK}|\}}|x_i| \leq \sqrt{2}\sum_{i\in\{|\mathcal{DK}|+1,\dots,|\mathcal{DK}|+\frac{N-1-3K}{2}\}}|x_i| \tag{10}$$

holds for *every* vector $\boldsymbol{x}$ in the null space of a uniform distributed M'-dimensional projection A' in $R^{|\mathcal{DK}|+\frac{N-1-3K}{2}}$, where $M' = |\mathcal{DK}|+\frac{N-1-3K}{2}-(N-M)$. This is because the null space of a uniform M-dimensional subspaces in R^n can be represented as $\{\boldsymbol{x} : \boldsymbol{x} = H\boldsymbol{z}, \boldsymbol{z} \in R^{N-M}\}$, where H is an $N\times(N-M)$ matrix whose elements are i.i.d. Gaussian random variables $\mathcal{N}(0,1)$. With H_l denoting the l-th row of H, $H_{i+1}-H_i$ is just a row vector with elements being i.i.d Gaussian random variables $\mathcal{N}(0,2)$. Noting $x_{i+1}-x_i =< H_{i+1}-H_i, \boldsymbol{z} >$, we can just think of $x_{i+1}-x_i$ as a $\sqrt{2}$ multiple of an element of a vector in a uniform M'-dimensional subspace in $R^{|\mathcal{DK}|+\frac{N-1-3K}{2}}$.

Now our problem reduces to determining for what values of K, with high probability the $RelaxedNULL$ condition (9) holds *simultaneously* for *every* gradient support set $\mathcal{K}$ (which determines $\mathcal{DK}$ and $\mathcal{KB}$). This falls exactly into the Grassmann angle framework [11,13,28] which can compute such K using the Grassmann angle tools from high dimensional convex polytope theory. For details, the reader can refer to [28], with the corresponding parameter C in [28] set as $\sqrt{2}$. Figure 1 plots the recoverable threshold $\frac{K}{M}$ as a function of the compression ration $\frac{M}{N}$ as $N\to\infty$.

2.4 Gradient Sparsity K Growing Linearly with Signal Dimension N

In this part, we consider the regime of interest where the sparsity of the signal gradient grows linearly with the problem dimension. The main result is summarized in the following theorem, showing that TV minimization can allow the gradient sparsity K to grow proportionally with signal dimension N.

Theorem 2.3. *Suppose that the measurement matrix A is an $M\times N$ matrix having i.i.d. standard zero mean Gaussian elements. For any constant $0<\alpha<1$, there exists a constant $\delta>0$ such that the following statement holds true, with overwhelming probability as $M\to\infty$, $N\to\infty$, and $\frac{M}{N}\to\alpha$.*

For all subsets $\mathcal{K}\subseteq\{1,2,...,N-1\}$ with cardinality $|\mathcal{K}|\leq\delta N$, and for every nonzero vector $\boldsymbol{x}$ in the null space of A (namely $A\boldsymbol{x}=0$, $\boldsymbol{x}\neq\boldsymbol{0}$),

$$\|(D\boldsymbol{x})_{\mathcal{K}}\|_1 < \|(D\boldsymbol{x})_{\mathcal{K}^c}\|_1, \tag{11}$$

where $\mathcal{K}^c = \{1,2,...,N-1\}\setminus K$.

To prove Theorem 2.3, we first prove a uniform lower bound for the TV norm in Subsection 2.4.1, and then utilize the lower bound to arrive at the conclusion in Subsection 2.4.2.

2.4.1 Uniform Lower Bound for Total Variation Norm

We consider the $(N-M)$-dimensional null space of the measurement matrix A. Recall that A has i.i.d. standard zero mean Gaussian elements. Equivalently, a basis for the null space of A can be represented by an $N \times (N-M)$ matrix H with i.i.d. standard zero mean Gaussian elements. To prove the null space property for successful signal recovery using TV minimization, we only need to prove the null space property holds for those vectors Hz, where $z \in \mathbb{R}^{N-M}$ with $\|z\|_2 = 1$.

To this end, we first establish the following claim.

Theorem 2.4. *With high probability as $N \to \infty$, uniformly for every $\mathbf{x} = Hz$ with $z \in \mathbb{R}^{N-M}$ and $\|z\|_2 = 1$,*

$$\|(D\boldsymbol{x})\|_1 \geq \gamma n,$$

where $\gamma > 0$ is a sufficiently small positive constant.

We divide the proof into three parts. In the first part (Subsection 2.4.1.1), we establish an upper bound uniformly true for every $\mathbf{x} = Hz$ with $z \in \mathbb{R}^{N-M}$ and $\|z\|_2 = 1$. In the second part (Subsection 2.4.1.2), assuming that a certain deviation bound holds true for the Total Variation norm (to be proven in Subsection 2.4.1.3), we establish Theorem 2.4 using the technique of ϵ-net. In the third part (Subsection 2.4.1.3), we prove the needed deviation bound for Total Variation norm.

2.4.1.1 Upper Bound for the Total Variation First of all, with high probability, as $N \to \infty$, for every $\mathbf{x} = Hz$ with $z \in \mathbb{R}^{N-M}$ and $\|z\|_2 = 1$,

$$\|\mathbf{x}\|_2 \leq C_1 \sqrt{N},$$

where C_1 is a constant as $N \to \infty$. We have used the deviation bound for the largest singular value of matrices with i.i.d. Gaussian elements [15].

Following this fact, we know

$$\|(D\boldsymbol{x})\|_1 \leq 2\|\mathbf{x}\|_1 \leq 2\sqrt{N}\|\mathbf{x}\|_2 \leq 2C_1 N.$$

2.4.1.2 Uniform Lower Bound on Total Variation through the ϵ-Net We cover the sphere $\{z| \ \|z\|_2 = 1\}$ with ϵ-net, where $\epsilon = C\gamma$, $C > 0$ and $\gamma > 0$ are constants we will choose later. ϵ-net is a finite set $V = \{v_1, ..., v_L\}$ on $\{z| \ \|z\|_2 = 1\}$ such that every point z from $\{z| \ \|z\|_2 = 1\}$, there is a $v_l \in V$ such that $\|z - v_l\|_2 \leq \epsilon$. The size of the ϵ-net can be taken no bigger than $(1 + \frac{2}{\epsilon})^{N-M}$.

From Subsection 2.4.1.3, we know that for every $\mathbf{x} = Hz$ generated by points from ϵ-net,

$$\|(D\boldsymbol{x})\|_1 \geq \gamma N,$$

where $\gamma > 0$ is a sufficiently small constant.

For any z such that $\|z\|_2 = 1$, there exists a point v_0 (we change the subscript numbering for V to index the order) in V such that $\|z - v_0\|_2 \triangleq \epsilon_1 \leq \epsilon$. Let z_1 denote $z - v_0$, then $\|z_1 - \epsilon_1 v_1\|_2 \triangleq \epsilon_2 \leq \epsilon_1 \epsilon \leq \epsilon^2$ for some v_1 in V. Repeating this process, we have $z = \sum_{j \geq 0} \epsilon_j v_j$, where $\epsilon_0 = 1$, $\epsilon_j \leq \epsilon^j$ and $v_j \in V$.

Then

$$\begin{aligned}
\sum_i |(D(Hz))_i| &= \sum_i |D(\sum_{j \geq 0} \epsilon_j H v_j)_i| \\
&\geq \sum_i (|D(Hv_0)_i| - \sum_{j \geq 1} \epsilon_j |D(Hv_j)_i|) \\
&\geq \sum_i |D(Hv_0)_i| - \sum_{j \geq 1} \epsilon^j \sum_i |D(Hv_j)_i| \\
&\geq \gamma N - \frac{\epsilon}{1-\epsilon} \times 2C_1 N,
\end{aligned}$$

where the first inequality follows from the triangle inequality, and the last inequality follows from the upper bound on the TV norm in Subsection 2.4.1.1.

We have just shown that, for for every $\mathbf{x} = Hz$ with $z \in \mathbb{R}^{N-M}$ and $\|z\|_2 = 1$,

$$\|(D\boldsymbol{x})\|_1 \geq \gamma N - \frac{\epsilon}{1-\epsilon} \times 2C_1 N.$$

For any arbitrary positive constant $\beta > 0$, we can always take $C > 0$ to be a sufficiently small constant (this does not affect the proof and conclusion in Subsection 2.4.1.3), such that

$$\|(D\boldsymbol{x})\|_1 \geq (1-\beta)\gamma N$$

holds true for every every $\mathbf{x} = Hz$ with $z \in \mathbb{R}^{N-M}$ and $\|z\|_2 = 1$.

2.4.1.3 Proving the deviation bound In this subsection, we prove that, for a constant $C > 0$, a sufficiently small constant $\gamma > 0$, and $\epsilon = C\gamma$, for every $\mathbf{x} = Hz$ generated by points from ϵ-net,

$$\|(D\boldsymbol{x})\|_1 \geq \gamma N,$$

with overwhelming probability as $N \to \infty$.

We claim, it is sufficient to prove, for a vector $\mathbf{x}$ with i.i.d. zero mean standard Gaussian random variables, as $N \to \infty$, with probability $e^{-N(\log(\frac{1}{\gamma})+C_2)}$, where $C_2 > 0$ is a constant,

$$\|(D\boldsymbol{x})\|_1 \leq \gamma N.$$

In fact, we recall that the size of the ϵ-net is at most $(1+\frac{2}{C\gamma})^{N-M}$, and notice that, for any point $z \in \mathbb{R}^{N-M}$ and $\|z\|_2 = 1$, the elements of Hz are i.i.d. standard zero mean Gaussian random variables. By a simple union bound, with probability at most

$$P = e^{(N-M)\log(1+\frac{2}{C\gamma})} \times e^{-N(\log(\frac{1}{\gamma})+C_2)},$$

there exist some point $\mathbf{x} = Hz$ with z from the ϵ-net, such that

$$\|(D\boldsymbol{x})\|_1 \leq \gamma N.$$

No matter what C we are looking at, if we take $\gamma > 0$ sufficiently small, this probability P converges to 0, as $N \to \infty$, $M \to \infty$, and $\frac{M}{N} \to \alpha$, where α is a constant. This means that with overwhelming probability, for all points from the ϵ-net,

$$\|(D\boldsymbol{x})\|_1 \geq \gamma N.$$

This leads to the conclusion in Subsection 2.4.1.2 that, with overwhelming probability, for every $\mathbf{x} = Hz$ with $z \in \mathbb{R}^{N-M}$ and $\|z\|_2 = 1$,

$$\|(D\boldsymbol{x})\|_1 \geq (1-\beta)\gamma N.$$

Now we focus on proving the following theorem about a sequence of i.i.d. zero mean Gaussian random variables of unit variance.

Theorem 2.5. *Suppose x_1, x_2, ..., and x_N are N independent random variables following the standard Gaussian distribution $\mathcal{N}(0,1)$. Then for all sufficiently small $\gamma > 0$, the probability*

$$P(\sum_{i=1}^{N-1} |x_{i+1} - x_i| \leq \gamma N) \leq e^{-(1-\mu)N(\log(\frac{1}{\gamma})+C_2+o(1))},$$

where both $\mu > 0$ and C_2 are constants, independent of γ and N; $o(1)$ goes to zero as $N \to \infty$; moreover, $\mu > 0$ can be made arbitrarily small.

Proof. Let us define $t_i = |x_{i+1} - x_i|$, $1 \leq i \leq N-1$. Suppose that

$$\|(D\boldsymbol{x})\|_1 \leq \gamma N,$$

then among N terms in $\|(D\boldsymbol{x})\|_1$, there must be at most $\frac{N}{T}$ terms which are larger than $T\gamma$ in magnitude, where T is any arbitrary positive number (say, $T = 1000$); namely, there must be at least $(1-\frac{1}{T})N$ terms that are no bigger than $T\gamma$.

Let $\mathcal{M}_i$ denote the event that $t_i \geq T\gamma$, and let $\mathcal{M}_i^c$ denote the complementary event that $t_i \leq T\gamma$. We further define the indicator function I_i, $1 \leq i \leq N$, as

$$I_i = \begin{cases} 0 & \text{if } M_i \text{ happens,} \\ 1 & \text{if } M_i^c \text{ happens.} \end{cases}$$

And let S be the set of i's such that $t_i \leq T\gamma$; and $S^c = \{1, 2, ..., N-1\} \setminus S$ be the complement of S.

Then the probability that M_i happens for and only for $i \in S$ is

$$P(I_1, I_2, ..., I_{N-1}) = \prod_{i=1}^{N-1} P(I_i | I_1, I_2, ..., I_{i-1}).$$

When $I_i = 0$, we simply upper bound $P(I_i|I_1, I_2, ..., I_{i-1})$ by 1; when $I_i = 1$, we claim that $P(I_i|I_1, I_2, ..., I_{i-1})$ is upper bounded by $\frac{1}{\sqrt{2\pi}}T\gamma$. In fact, in the Gaussian Markov chain $x_{i+1} - x_i$, no matter what values x_1,x_2, ...x_{i-1} take, the probability of having a magnitude $|x_i - x_{i-1}|$ no larger than than $T\gamma$ is maximized when x_{i-1} is equal to 0. This leads to

$$P(I_1, I_2, ..., I_{N-1}) \leq (\frac{1}{\sqrt{2\pi}}T\gamma)^{|S|}.$$

Since $|S| \geq (1-\frac{1}{T})N$, we have

$$P(I_1, I_2, ..., I_{N-1}) \leq (\frac{1}{\sqrt{2\pi}}T\gamma)^{(1-\frac{1}{T})N}.$$

We have at most $\binom{N-1}{|S|}$ possibility for the set S with cardinality $|S|$. So the probability

$$P_1 \triangleq P(\sum_{i=1}^{N-1} |x_{i+1} - x_i| \leq \gamma N) \leq \sum_{j=(1-\frac{1}{T})N}^{N-1} \binom{N-1}{j} (\frac{1}{\sqrt{2\pi}}T\gamma)^j.$$

From Stirling's formula and notice the $j = (1-\frac{1}{T})N$ is the biggest term in the upper bound of P_1, when γ is sufficient small such that $(\frac{1}{\sqrt{2\pi}}T\gamma)$ is smaller than $1-\frac{1}{T}$, we have

$$\log(P_1)/N \to \left(H(\frac{1}{T}) + (1-\frac{1}{T})\log(\frac{1}{\sqrt{2\pi}}T\gamma)\right) + o(1),$$

where $H(p) = p\log(\frac{1}{p}) + (1-p)\log(\frac{1}{1-p})$ is the entropy function, and $o(1)$ is a term that goes to 0 as $N \to \infty$.

Since we can pick an arbitrarily big constant T, we have the theorem statement by simply taking $\mu = \frac{1}{T}$. $\square$

We remark that Theorem 2.5 eventually leads to the proof of Theorem 2.4.

2.4.2 Upper Bound on the Partial Total Variation Norm

In this section, we prove the following theorem.

Theorem 2.6. *Suppose a matrix H is an $N \times (N-M)$ matrix having i.i.d. standard zero mean Gaussian elements. For any constant $0 < \alpha < 1$ and any positive constant $\gamma > 0$, there exists a constant $\delta > 0$ such that the following statement holds true, with overwhelming probability as $M \to \infty$, $N \to \infty$, and $\frac{M}{N} \to \alpha$.*

For all subsets $\mathcal{K} \subseteq \{1, 2, ..., N-1\}$ with cardinality $|\mathcal{K}| \leq \delta N$, and for every $\mathbf{x} = Hz$ with $z \in \mathbb{R}^{N-M}$ and $\|z\|_2 = 1$,

$$\|(D\boldsymbol{x})_{\mathcal{K}}\|_1 < \frac{1}{2}\gamma N. \tag{12}$$

We notice that $\|(D\boldsymbol{x})_{\mathcal{K}}\|_1 \leq 2 \sum_{j \in \mathcal{M}} |x_j|$, where $\mathcal{M}$ is the set of indices j's that x_j is involved in the expression $\|(D\boldsymbol{x})_{\mathcal{K}}\|_1$. Because of this, and the fact that the cardinality $|\mathcal{M}| \leq 2|\mathcal{K}|$, we can use the same methodology in [27] to prove Theorem 2.6, based on the uniform lower bound result we have from Theorem 2.4.

3 Extension to Multidimensional signals

In this section, we extend our results to d-dimensional ($d \geq 2$, for example $d = 2$ for image and $d = 3$ for videos) signal vectors. We get results that are comparable to those in [19, 20]. In particular, let $\boldsymbol{X} \in \mathbb{R}^{N^d}$ be a multi-indexed vector that is from a d-dimensional signal. Let $A \in \mathbb{R}^{M \times N^d}$ be a measurement matrix whose elements are i.i.d. Gaussian random variables, and $\boldsymbol{Y} = A\boldsymbol{X}$ be its corresponding measurements of $\boldsymbol{X}$. Define $D\boldsymbol{X}$ be the discrete gradient of $\boldsymbol{X}$. Assume that $D\boldsymbol{X}$ contains at most K nonzero entries. In order to recover $\boldsymbol{X}$, similar to (2), we solve the following minimization

$$\min_{\boldsymbol{X}} \ \|D\boldsymbol{X}\|_1, \qquad \text{subject to} \quad \boldsymbol{Y} = A\boldsymbol{X}. \tag{13}$$

In the remaining of this section, we prove that the unique solution of (13) is exactly the original $\boldsymbol{X}$ with high probability, as long as

$$M \geq \begin{cases} C_1 K \log_2^2 N \ln N & \text{if } d = 2 \\ C_2 K \ln N & \text{if } d > 2. \end{cases}$$

where $C_1 > 0$ and $C_2 > 0$ are two constants depending on d. Note that $\|D\boldsymbol{X}\|_1$ in (13) is the anisotropic TV. Our proof can be generalized to isotropic TV without too much difficulty.

Similar to Theorem 2.1, a sufficient condition for the original $\boldsymbol{X}$ being the unique solution of (13) is the null space condition (3). Different from 1-dimensional case, this null space condition is only a sufficient condition for higher dimensional signals. Then, using the escape through the mesh theorem, this null space condition holds true with high probability if the Gaussian width satisfies $w(\mathcal{S}_d) < \sqrt{M} - \frac{1}{2\sqrt{M}}$, where

$$\mathcal{S}_d = \{\boldsymbol{X} \in \mathbb{R}^{N^d} \ : \ \|\boldsymbol{X}\|_2 = 1, \quad \text{and} \quad \|(D\boldsymbol{X})_{\mathcal{K}}\|_1 \geq \|(D\boldsymbol{X})_{\mathcal{K}^c}\|_1 \ \exists \mathcal{K} \subset \{1, \ldots, N\}^d \times \{1, \ldots, d\} \text{ s.t. } |\mathcal{K}| \leq K\}.$$

Given any vector $\boldsymbol{X} \in \mathcal{S}_d$, we have

$$\begin{aligned} \|D\boldsymbol{X}\|_1 &= \|(D\boldsymbol{X})_{\mathcal{K}^c}\|_1 + \|(D\boldsymbol{X})_{\mathcal{K}}\|_1 \leq 2\|(D\boldsymbol{X})_{\mathcal{K}}\|_1 \leq 2\sqrt{K}\|(D\boldsymbol{X})_{\mathcal{K}}\|_2 \\ &\leq 2\sqrt{K}\|D\boldsymbol{X}\|_2 \leq 4\sqrt{d}\sqrt{K}\|\boldsymbol{X}\|_2 \leq 4\sqrt{d}\sqrt{K}. \end{aligned}$$

We have used the fact that $\|D\boldsymbol{X}\|_2 \leq 2\sqrt{d}\|\boldsymbol{X}\|_2$. Therefore,

$$\mathcal{S}_d \subset \widetilde{\mathcal{S}}_d := \{\boldsymbol{X} \in \mathbb{R}^{N^d} \ : \ \|\boldsymbol{X}\|_2 \leq 1, \quad \|D\boldsymbol{X}\|_1 \leq 4\sqrt{d}\sqrt{K}\}.$$

In the following, we estimate the Gaussian width of $\widetilde{\mathcal{S}}_d$. Similar to 1-dimensional signal, we consider only the case where $N = 2^L$. For any $\boldsymbol{X} \in \widetilde{\mathcal{S}}_d$, we decompose $\boldsymbol{X}$ according to Haar wavelet transform for d-dimensional vector as

$$\boldsymbol{X} = \sum_{\ell=1}^{L} \sum_{\boldsymbol{i} \in \{0,1\}^d \setminus \boldsymbol{0}} \hat{\boldsymbol{Z}}^{(\ell, \boldsymbol{i})} + \hat{\boldsymbol{Y}}^{(L)}, \tag{14}$$

where

$$\hat{\boldsymbol{Z}}^{(\ell,\boldsymbol{i})} = \boldsymbol{Z}^{(\ell,\boldsymbol{i})} \otimes \boldsymbol{H}^{(\boldsymbol{i})} \otimes \mathbf{1}_{2^{\ell-1}}, \quad \boldsymbol{Z}^{(\ell,\boldsymbol{i})} \in \mathbb{R}^{(N/2^\ell)^d}$$

and

$$\hat{\boldsymbol{Y}}^{(L)} = \boldsymbol{Y}^{(L)} \otimes \mathbf{1}_N, \quad \boldsymbol{Y}^{(L)} \in \mathbb{R}.$$

Here $\mathbf{1}_n \in \mathbb{R}^{n^d}$ is the d-dimensional vector whose entries are all 1, and $\otimes$ is the Kronecker product, i.e., $\boldsymbol{A} \otimes \boldsymbol{B}$ is the block d-dimensional matrix whose $(j_1, j_2, \ldots, j_d)$ block is $A_{j_1 j_2 \ldots j_d}\boldsymbol{B}$. Moreover, $\boldsymbol{H}^{(\boldsymbol{i})} \in \mathbb{R}^{2^d}$ with $\boldsymbol{i} = (i_1, i_2, \ldots, i_d)$ is the (scaled) Haar filter defined by

$$H^{(\boldsymbol{i})}_{j_1 j_2 \ldots j_d} = \prod_{k=1}^{d} h^{(i_k)}_{j_k}, \quad \text{with} \quad \boldsymbol{h}^{(0)} = [1\ 1],\ \boldsymbol{h}^{(1)} = [1\ -1].$$

In particular, we have $\boldsymbol{H}^{(0)} = \mathbf{1}_2$.

The decomposition (14) is done recursively as follows. We first decompose $\boldsymbol{X} := \hat{\boldsymbol{Y}}^{(0)} = \hat{\boldsymbol{Y}}^{(1)} + \sum_{\boldsymbol{i}\in\{0,1\}^d\backslash\mathbf{0}} \hat{\boldsymbol{Z}}^{(1,\boldsymbol{i})}$, where

$$\hat{\boldsymbol{Y}}^{(1)} = \boldsymbol{Y}^{(1)} \otimes \mathbf{1}_2, \qquad Y^{(1)}_{\boldsymbol{k}} = \frac{\sum_{\boldsymbol{j}\in\{0,1\}^d} H^{(\mathbf{0})}_{\boldsymbol{j}} X_{2\boldsymbol{k}-\boldsymbol{j}}}{2^d} = \frac{\sum_{\boldsymbol{j}\in\{0,1\}^d} X_{2\boldsymbol{k}-\boldsymbol{j}}}{2^d},$$

and

$$\hat{\boldsymbol{Z}}^{(1,\boldsymbol{i})} = \boldsymbol{Z}^{(1,\boldsymbol{i})} \otimes \boldsymbol{H}^{(\boldsymbol{i})}, \qquad Z^{(1,\boldsymbol{i})}_{\boldsymbol{k}} = \frac{\sum_{\boldsymbol{j}\in\{0,1\}^d} H^{(\boldsymbol{i})}_{\boldsymbol{j}} X_{2\boldsymbol{k}-\boldsymbol{j}}}{2^d}.$$

One can check that $\hat{\boldsymbol{Y}}^{(0)} = \hat{\boldsymbol{Y}}^{(1)} + \sum_{\boldsymbol{i}\in\{0,1\}^d\backslash\mathbf{0}} \hat{\boldsymbol{Z}}^{(1,\boldsymbol{i})}$. Furthermore, it can be easily shown that this decomposition is an orthogonal decomposition. Then, we further decompose

$$\hat{\boldsymbol{Y}}^{(1)} = \hat{\boldsymbol{Y}}^{(2)} + \sum_{\boldsymbol{i}\in\{0,1\}^d\backslash\mathbf{0}} \hat{\boldsymbol{Z}}^{(2,\boldsymbol{i})} \tag{15}$$

where

$$\hat{\boldsymbol{Y}}^{(2)} = \boldsymbol{Y}^{(2)} \otimes \mathbf{1}_4, \qquad Y^{(2)}_{\boldsymbol{k}} = \frac{\sum_{\boldsymbol{j}\in\{0,1\}^d} H^{(\boldsymbol{i})}_{\boldsymbol{j}} Y^{(1)}_{2\boldsymbol{k}-\boldsymbol{j}}}{2^d} = \frac{\sum_{\boldsymbol{j}\in\{0,1\}^d} Y^{(1)}_{2\boldsymbol{k}-\boldsymbol{j}}}{2^d},$$

and

$$\hat{\boldsymbol{Z}}^{(2,\boldsymbol{i})} = \boldsymbol{Z}^{(2,\boldsymbol{i})} \otimes \boldsymbol{H}^{(\boldsymbol{i})} \otimes \mathbf{1}_2, \qquad Z^{(2,\boldsymbol{i})}_{\boldsymbol{k}} = \frac{\sum_{\boldsymbol{j}\in\{0,1\}^d} H^{(\boldsymbol{i})}_{\boldsymbol{j}} Y^{(1)}_{2\boldsymbol{k}-\boldsymbol{j}}}{2^d}.$$

Again, one can check (15) holds true and is an orthogonal decomposition. Generally, at level ℓ, we have that

$$\hat{\boldsymbol{Y}}^{(\ell)} = \boldsymbol{Y}^{(\ell)} \otimes \mathbf{1}_{2^\ell}, \qquad \boldsymbol{Y}^{(\ell)} \in \mathbb{R}^{(N/2^\ell)^d},$$

and we decompose it as

$$\hat{\boldsymbol{Y}}^{(\ell)} = \hat{\boldsymbol{Y}}^{(\ell+1)} + \sum_{\boldsymbol{i}\in\{0,1\}^d\backslash\mathbf{0}} \hat{\boldsymbol{Z}}^{(\ell+1,\boldsymbol{i})}, \tag{16}$$

where

$$\hat{\boldsymbol{Y}}^{(\ell+1)} = \boldsymbol{Y}^{(\ell+1)} \otimes \mathbf{1}_{2^{\ell+1}}, \qquad Y^{(\ell+1)}_{\boldsymbol{k}} = \frac{\sum_{\boldsymbol{j}\in\{0,1\}^d} H^{(\mathbf{0})}_{\boldsymbol{j}} Y^{(\ell)}_{2\boldsymbol{k}-\boldsymbol{j}}}{2^d} = \frac{\sum_{\boldsymbol{j}\in\{0,1\}^d} Y^{(\ell)}_{2\boldsymbol{k}-\boldsymbol{j}}}{2^d},$$

and

$$\hat{\boldsymbol{Z}}^{(\ell+1,\boldsymbol{i})} = \boldsymbol{Z}^{(\ell+1,\boldsymbol{i})} \otimes \boldsymbol{H}^{(\boldsymbol{i})} \otimes \mathbf{1}_{2^\ell}, \qquad Z^{(\ell+1,\boldsymbol{i})}_{\boldsymbol{k}} = \frac{\sum_{\boldsymbol{j}\in\{0,1\}^d} H^{(\boldsymbol{i})}_{\boldsymbol{j}} Y^{(\ell)}_{2\boldsymbol{k}-\boldsymbol{j}}}{2^d}.$$

The decomposition (14) has the following properties.

- Obviously, components in decomposition (14) are orthogonal to each others. Consequently,

$$\|\boldsymbol{X}\|_2^2 = \sum_{\ell=1}^{L} \sum_{\boldsymbol{i}\in\{0,1\}^d\setminus\boldsymbol{0}} \|\hat{\boldsymbol{Z}}^{(\ell,\boldsymbol{i})}\|_2^2 + \|\hat{\boldsymbol{Y}}^{(L)}\|_2^2 = \sum_{\ell=1}^{L}\left(2^{d\ell} \sum_{\boldsymbol{i}\in\{0,1\}^d\setminus\boldsymbol{0}} \|\hat{\boldsymbol{Z}}^{(\ell,\boldsymbol{i})}\|_2^2\right) + 2^{dL}\|\hat{\boldsymbol{Y}}^{(L)}\|_2^2.$$

  Since $\boldsymbol{X}\in\tilde{\mathcal{S}}$ implies $\|\boldsymbol{X}\|_2^2 \le 1$, we have

$$\sum_{\ell=1}^{L}\left(2^{d\ell} \sum_{\boldsymbol{i}\in\{0,1\}^d\setminus\boldsymbol{0}} \|\hat{\boldsymbol{Z}}^{(\ell,\boldsymbol{i})}\|_2^2\right) + 2^{dL}\|\hat{\boldsymbol{Y}}^{(L)}\|_2^2 \le 1 \tag{17}$$

- It can be shown that

$$\|D\hat{\boldsymbol{Y}}^{(\ell)}\|_1 \le \|D\hat{\boldsymbol{Y}}^{(\ell-1)}\|_1 \tag{18}$$

  and, consequently,

$$\|D\hat{\boldsymbol{Y}}^{(\ell)}\|_{(K)}^* \le \|D\boldsymbol{X}\|_{(K)}^* \le 4\sqrt{d}\sqrt{K}. \tag{19}$$

  Let D_i be the difference matrix along the i-th dimension. Then, similar to the 1-D case, one can show that $\|D_i\boldsymbol{Y}^{(\ell)}\|_1 \le 2^{d-1}\cdot\frac{2}{2^d}\|D_i\boldsymbol{Y}^{(\ell-1)}\|_1 = \|D_i\boldsymbol{Y}^{(\ell-1)}\|_1$. Summing over i yields (18).

- Furthermore, for any vector $\boldsymbol{G}$, we have

$$\sum_{\boldsymbol{i}\in\{0,1\}^d\setminus\boldsymbol{0}} \langle \boldsymbol{G}, \hat{\boldsymbol{Z}}^{(\ell,\boldsymbol{i})}\rangle \le \frac{2^d(2^d-1)}{2^{d-1}}\frac{4\sqrt{d}\sqrt{K}}{2^{(\ell-1)(d-1)}}\|\tilde{\boldsymbol{G}}^{(\ell-1)}\|_\infty = 8\sqrt{d}(2^d-1)\frac{\sqrt{K}}{2^{(\ell-1)(d-1)}}\|\tilde{\boldsymbol{G}}^{(\ell-1)}\|_\infty,$$

  where $\tilde{\boldsymbol{G}}^{(\ell-1)} \in \mathbb{R}^{(N/2^{\ell-1})^d}$ is a d-dimensional signal whose $\boldsymbol{i}$-th entry is the sum of the entries of $\boldsymbol{G}$ on the $\boldsymbol{i}$-th block of size $2^{\ell-1}\times 2^{\ell-1}$. For simplicity, we prove it for $d=2$. The remaining case can be shown analogously. When $d=2$, we have the four filters are

$$\boldsymbol{H}^{(0,0)} = \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix},\quad \boldsymbol{H}^{(1,0)} = \begin{bmatrix}1 & -1\\ 1 & -1\end{bmatrix},\quad \boldsymbol{H}^{(0,1)} = \begin{bmatrix}1 & 1\\ -1 & -1\end{bmatrix},\quad \boldsymbol{H}^{(1,1)} = \begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}.$$

  Let D_1 and D_2 be finite difference along the horizontal and vertical direction respectively. Then it is easy to check that

$$\begin{aligned}&\|D(a_{00}\boldsymbol{H}^{(0,0)}\otimes\mathbf{1}_{2^{\ell-1}} + a_{10}\boldsymbol{H}^{(1,0)}\otimes\mathbf{1}_{2^{\ell-1}} + a_{01}\boldsymbol{H}^{(0,1)}\otimes\mathbf{1}_{2^{\ell-1}} + a_{11}\boldsymbol{H}^{(1,1)}\otimes\mathbf{1}_{2^{\ell-1}})\|_1\\ &= \|D_1(a_{00}\boldsymbol{H}^{(0,0)}\otimes\mathbf{1}_{2^{\ell-1}} + a_{10}\boldsymbol{H}^{(1,0)}\otimes\mathbf{1}_{2^{\ell-1}} + a_{01}\boldsymbol{H}^{(0,1)}\otimes\mathbf{1}_{2^{\ell-1}} + a_{11}\boldsymbol{H}^{(1,1)}\otimes\mathbf{1}_{2^{\ell-1}})\|_1\\ &\quad + \|D_2(a_{00}\boldsymbol{H}^{(0,0)}\otimes\mathbf{1}_{2^{\ell-1}} + a_{10}\boldsymbol{H}^{(1,0)}\otimes\mathbf{1}_{2^{\ell-1}} + a_{01}\boldsymbol{H}^{(0,1)}\otimes\mathbf{1}_{2^{\ell-1}} + a_{11}\boldsymbol{H}^{(1,1)}\otimes\mathbf{1}_{2^{\ell-1}})\|_1\\ &= 2^{\ell-1}(|a_{01}+a_{11}| + |a_{01}-a_{11}| + |a_{10}+a_{11}| + |a_{10}-a_{11}|).\end{aligned}$$

  Therefore,

$$\begin{aligned}&2^{\ell-1}(\|\boldsymbol{Z}^{(\ell,01)}+\boldsymbol{Z}^{(\ell,11)}\|_1 + \|\boldsymbol{Z}^{(\ell,01)}-\boldsymbol{Z}^{(\ell,11)}\|_1 + \|\boldsymbol{Z}^{(\ell,10)}+\boldsymbol{Z}^{(\ell,11)}\|_1 + \|\boldsymbol{Z}^{(\ell,10)}-\boldsymbol{Z}^{(\ell,11)}\|_1)\\ &\qquad \le \|D(\hat{\boldsymbol{Y}}^{(\ell)} + \hat{\boldsymbol{Z}}^{(\ell,10)} + \hat{\boldsymbol{Z}}^{(\ell,01)} + \hat{\boldsymbol{Z}}^{(\ell,11)})\|_1 = \|D\hat{\boldsymbol{Y}}^{(\ell-1)}\|_1 \le 4\sqrt{2}\sqrt{K}.\end{aligned}$$

  Furthermore, if we let $\tilde{\boldsymbol{G}}_{oo}^{(\ell-1)}$ be a down sample of $\tilde{\boldsymbol{G}}^{(\ell-1)}$ on odd-odd indices and similarly $\tilde{\boldsymbol{G}}_{oe}^{(\ell-1)}$, $\tilde{\boldsymbol{G}}_{eo}^{(\ell-1)}$, and $\tilde{\boldsymbol{G}}_{ee}^{(\ell-1)}$, then

$$\begin{aligned}&\langle \boldsymbol{G}, \hat{\boldsymbol{Z}}^{(\ell,10)} + \hat{\boldsymbol{Z}}^{(\ell,01)} + \hat{\boldsymbol{Z}}^{(\ell,11)}\rangle = \langle \tilde{\boldsymbol{G}}^{(\ell-1)}, \boldsymbol{Z}^{(\ell,10)}\otimes\boldsymbol{H}^{(1,0)} + \boldsymbol{Z}^{(\ell,01)}\otimes\boldsymbol{H}^{(0,1)} + \boldsymbol{Z}^{(\ell,11)}\otimes\boldsymbol{H}^{(1,1)}\rangle\\ =&\langle \tilde{\boldsymbol{G}}_{oo}^{(\ell-1)}, \boldsymbol{Z}^{(\ell,01)} + \boldsymbol{Z}^{(\ell,10)} + \boldsymbol{Z}^{(\ell,11)}\rangle + \langle \tilde{\boldsymbol{G}}_{oe}^{(\ell-1)}, -\boldsymbol{Z}^{(\ell,01)} + \boldsymbol{Z}^{(\ell,10)} - \boldsymbol{Z}^{(\ell,11)}\rangle\\ &+ \langle \tilde{\boldsymbol{G}}_{eo}^{(\ell-1)}, \boldsymbol{Z}^{(\ell,01)} - \boldsymbol{Z}^{(\ell,10)} - \boldsymbol{Z}^{(\ell,11)}\rangle + \langle \tilde{\boldsymbol{G}}_{ee}^{(\ell-1)}, -\boldsymbol{Z}^{(\ell,01)} - \boldsymbol{Z}^{(\ell,10)} + \boldsymbol{Z}^{(\ell,11)}\rangle\\ \le&\|\tilde{\boldsymbol{G}}_{oo}^{(\ell-1)}\|_\infty\|\boldsymbol{Z}^{(\ell,01)} + \boldsymbol{Z}^{(\ell,10)} + \boldsymbol{Z}^{(\ell,11)}\|_1 + \|\tilde{\boldsymbol{G}}_{oe}^{(\ell-1)}\|_\infty\|-\boldsymbol{Z}^{(\ell,01)} + \boldsymbol{Z}^{(\ell,10)} - \boldsymbol{Z}^{(\ell,11)}\|_1\\ &+ \|\tilde{\boldsymbol{G}}_{eo}^{(\ell-1)}\|_\infty\|\boldsymbol{Z}^{(\ell,01)} - \boldsymbol{Z}^{(\ell,10)} - \boldsymbol{Z}^{(\ell,11)}\|_1 + \|\tilde{\boldsymbol{G}}_{ee}^{(\ell-1)}\|_\infty\|-\boldsymbol{Z}^{(\ell,01)} - \boldsymbol{Z}^{(\ell,10)} + \boldsymbol{Z}^{(\ell,11)}\|_1\end{aligned}$$

Since $\boldsymbol{Z}^{(\ell,01)} = \frac{1}{2}((\boldsymbol{Z}^{(\ell,01)} + \boldsymbol{Z}^{(\ell,11)}) + (\boldsymbol{Z}^{(\ell,01)} - \boldsymbol{Z}^{(\ell,11)}))$, $\boldsymbol{Z}^{(\ell,10)} = \frac{1}{2}((\boldsymbol{Z}^{(\ell,10)} + \boldsymbol{Z}^{(\ell,11)}) + (\boldsymbol{Z}^{(\ell,10)} - \boldsymbol{Z}^{(\ell,11)}))$, and $\boldsymbol{Z}^{(\ell,11)} = \frac{1}{2}((\boldsymbol{Z}^{(\ell,01)} + \boldsymbol{Z}^{(\ell,11)}) - (\boldsymbol{Z}^{(\ell,01)} - \boldsymbol{Z}^{(\ell,11)}))$, we have

$$\begin{aligned}
&\|\boldsymbol{Z}^{(\ell,01)} + \boldsymbol{Z}^{(\ell,10)} + \boldsymbol{Z}^{(\ell,11)}\|_1 \le \frac{1}{2}(\|\boldsymbol{Z}^{(\ell,01)} + \boldsymbol{Z}^{(\ell,11)}\|_1 + \|\boldsymbol{Z}^{(\ell,01)} - \boldsymbol{Z}^{(\ell,11)}\|_1) \\
&+ \frac{1}{2}(\|\boldsymbol{Z}^{(\ell,10)} + \boldsymbol{Z}^{(\ell,11)}\|_1 + \|\boldsymbol{Z}^{(\ell,10)} - \boldsymbol{Z}^{(\ell,11)}\|_1) + \frac{1}{2}(\|\boldsymbol{Z}^{(\ell,01)} + \boldsymbol{Z}^{(\ell,11)}\|_1 + \|\boldsymbol{Z}^{(\ell,01)} - \boldsymbol{Z}^{(\ell,11)}\|_1) \\
&\le \frac{3}{2}\frac{4\sqrt{2}\sqrt{K}}{2^{\ell-1}}
\end{aligned}$$

and similarly,

$$\|\boldsymbol{Z}^{(\ell,01)} - \boldsymbol{Z}^{(\ell,10)} - \boldsymbol{Z}^{(\ell,11)}\|_1, \|-\boldsymbol{Z}^{(\ell,01)} - \boldsymbol{Z}^{(\ell,10)} + \boldsymbol{Z}^{(\ell,11)}\|_1, \|-\boldsymbol{Z}^{(\ell,01)} + \boldsymbol{Z}^{(\ell,10)} - \boldsymbol{Z}^{(\ell,11)}\|_1 \le \frac{3}{2}\frac{4\sqrt{2}\sqrt{K}}{2^{\ell-1}}$$

Therefore,

$$\langle \boldsymbol{G}, \hat{\boldsymbol{Z}}^{(\ell,10)} + \hat{\boldsymbol{Z}}^{(\ell,01)} + \hat{\boldsymbol{Z}}^{(\ell,11)}\rangle \le \frac{3\cdot 4}{2}\frac{4\sqrt{2}\sqrt{K}}{2^{\ell-1}}\|\tilde{\boldsymbol{G}}^{(\ell-1)}\|_\infty = 8\cdot\sqrt{2}\cdot 3\frac{\sqrt{K}}{2^{\ell-1}}\|\tilde{\boldsymbol{G}}^{(\ell-1)}\|_\infty$$

Now we are ready to estimate the Gaussian width of $\tilde{\mathcal{S}}_d$. Let $\boldsymbol{G}$ be a vector whose entries are i.i.d. Gaussian random variables with mean 0 and variance 1. The same argument in one dimensional cases leads to

$$E(\|\tilde{\boldsymbol{G}}^{(\ell)}\|_\infty) \le \sqrt{2^{d\ell}2\ln\left(e^{1/2}N^d/2^{d\ell}\right)}$$

which implies

$$\begin{aligned}
E\left(\sum_{\boldsymbol{i}\in\{0,1\}^d\setminus\boldsymbol{0}}\langle \boldsymbol{G}, \hat{\boldsymbol{Z}}^{(\ell,\boldsymbol{i})}\rangle\right) &\le 8\sqrt{d}(2^d-1)\frac{\sqrt{K}}{2^{(\ell-1)(d-1)}}E(\|\tilde{\boldsymbol{G}}^{(\ell-1)}\|_\infty) \\
&\le 8\sqrt{d}(2^d-1)\frac{\sqrt{K}}{2^{(\ell-1)(d-1)}}\sqrt{2^{d(\ell-1)}2\ln\left(e^{1/2}N^d/2^{d(\ell-1)}\right)} \\
&\le 8\sqrt{d}(2^d-1)\frac{\sqrt{K}}{2^{(\ell-1)(d-1)}}\sqrt{2^{d(\ell-1)}2\ln\left(e^{1/2}N^d\right)} \\
&= 8\sqrt{d}(2^d-1)\sqrt{K}2^{(\ell-1)(1-\frac{d}{2})}\sqrt{2\ln\left(e^{1/2}N^d\right)}.
\end{aligned}$$

Moreover,

$$E\left(\langle \boldsymbol{G}, \hat{\boldsymbol{Y}}^{(L)}\rangle\right) = E\left(Y^{(L)}\tilde{\boldsymbol{G}}^{(L)}\right) \le |Y^{(L)}|E\left(\|\tilde{\boldsymbol{G}}^{(L)}\|_\infty\right) \le \sqrt{\frac{1}{2^{dL}}}\sqrt{2^{dL}2\ln\left(e^{1/2}N^d/2^{dL}\right)} = \sqrt{3}.$$

Therefore,

$$\begin{aligned}
E(\langle \boldsymbol{G}, \boldsymbol{X}\rangle) &= \sum_{\ell=1}^{L-1} E\left(\sum_{\boldsymbol{i}\in\{0,1\}^d\setminus\boldsymbol{0}}\langle \boldsymbol{G}, \hat{\boldsymbol{Z}}^{(\ell,\boldsymbol{i})}\rangle\right) + E\left(\langle \boldsymbol{G}, \hat{\boldsymbol{Y}}^{(L)}\rangle\right) \\
&= 8\sqrt{d}(2^d-1)\sqrt{K}\sqrt{2\ln\left(e^{3/2}N^d\right)}\sum_{\ell=1}^{L-1}2^{(\ell-1)(1-\frac{d}{2})} + \sqrt{3} \\
&\le \begin{cases} 8\sqrt{d}(2^d-1)\sqrt{K}\sqrt{2\ln\left(e^{1/2}N^d\right)}\log_2 N + \sqrt{3} & \text{if } d=2, \\ 8\sqrt{d}(2^d-1)\sqrt{K}\sqrt{2\ln\left(e^{1/2}N^d\right)}\frac{2^{1-\frac{d}{2}}}{1-2^{1-\frac{d}{2}}} + \sqrt{3} & \text{if } d>2 \end{cases}
\end{aligned}$$

We require the Gaussian width is about $\sqrt{M}$, where M is the number of measurement. So, we have

$$M \sim \begin{cases} K\log_2^2 N \ln N & \text{if } d=2 \\ K\ln N & \text{if } d>2. \end{cases}$$

4 Conclusion

In this paper,we establish the proof for the performance guarantee of total variation (TV) minimization in recovering *one-dimensional* signal with sparse gradient support. This partially answers the open problem of proving the fidelity of total variation minimization in such a setting [20]. We also extend our results to TV minimization for multidimensional signals. Recoverable sparsity thresholds of TV minimization are explicitly computed for 1-dimensional signal by using the Grassmann angle framework.

Our current results work only for the Gaussian ensemble of measurement matrices. One future direction is to extend our results to general deterministic and random measurement matrices, such as partial Fourier matrices, and random Bernoulli matrices. Another direction we would like to pursue is to tighten our bounds for 1-dimensional signal vector. For multidimensional signals, we conjecture that for Gaussian measurement operators, when the number of measurements is proportional to the problem dimension N^d, the recoverable sparsity of gradient support, by the TV minimization, can also grow proportionally with N^d. We are also interested in working towards tightening our results in this direction. The almost Euclidean property of subspaces [17, 29, 30] can be further used to extend our results to proving the stability of TV minimization for signals with approximately sparse gradients, under noisy measurements.

References

[1] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, *Image restoration: Total variation, wavelet frames, and beyond*, J. Amer. Math. Soc., 25 (2012), pp. 1033–1089.

[2] E. J. Candès, *The restricted isometry property and its implications for compressed sensing*, C. R. Math. Acad. Sci. Paris, 346 (2008), pp. 589–592.

[3] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, *Compressed sensing with coherent and redundant dictionaries*, Appl. Comput. Harmon. Anal., 31 (2011), pp. 59–73.

[4] E. J. Candès and P. A. Randall, *Highly robust error correction by convex programming*, IEEE Trans. Inform. Theory, 54 (2008), pp. 2829–2840.

[5] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

[6] E. J. Candes and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4203–4215.

[7] A. Chambolle and J. Darbon, *On total variation minimization and surface evolution using parametric maximum flows*, International Journal of Computer Vision, 84 (2009), pp. 288–307.

[8] A. d'Aspremont and L. El Ghaoui, *Testing the nullspace property using semidefinite programming*, Math. Program., 127 (2011), pp. 123–144.

[9] D. Donoho, I. Johnstone, and A. Montanari, *Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising*, http://arxiv.org/abs/1111.1041, (2013).

[10] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[11] ———, *High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension*, Discrete Comput. Geom., 35 (2006), pp. 617–652.

[12] D. L. Donoho, A. Maleki, and A. Montanari, *The noise-sensitivity phase transition in compressed sensing*, IEEE Trans. Inform. Theory, 57 (2011), pp. 6920–6941.

[13] D. L. Donoho and J. Tanner, *Neighborliness of randomly projected simplices in high dimensions*, Proc. Natl. Acad. Sci. USA, 102 (2005), pp. 9452–9457 (electronic).

[14] A. Y. Garnaev and E. D. Gluskin, *The widths of a Euclidean ball*, Dokl. Akad. Nauk SSSR, 277 (1984), pp. 1048–1052.

[15] S. Geman, *A limit theorem for the norm of random matrices*, Ann. Probab., 8 (1980), pp. 252–261.

[16] Y. Gordon, *On Milman's inequality and random subspaces which escape through a mesh in* $\mathbf{R}^n$, in Geometric aspects of functional analysis (1986/87), vol. 1317 of Lecture Notes in Math., Springer, Berlin, 1988, pp. 84–106.

[17] B. S. Kašin, *The widths of certain finite-dimensional sets and classes of smooth functions*, Izv. Akad. Nauk SSSR Ser. Mat., 41 (1977), pp. 334–351, 478.

[18] S. L. Keeling, *Total variation based convex filters for medical imaging*, Appl. Math. Comput., 139 (2003), pp. 101–119.

[19] D. Needell and R. Ward, *Stable image reconstruction using total variation minimization*, arXiv preprint arXiv:1202.6429, (2012).

[20] ———, *Total variation minimization for stable multidimensional signal recovery*, arXiv preprint arXiv:1210.3098, (2012).

[21] H. Rauhut, K. Schnass, and P. Vandergheynst, *Compressed sensing and redundant dictionaries*, IEEE Transactions on Information Theory, 54 (2008), pp. 2210–2219.

[22] M. Rudelson and R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, Comm. Pure Appl. Math., 61 (2008), pp. 1025–1045.

[23] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[24] E. Sidky and X. Pan, *Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization*, Physics in Medicine and Biology, 53 (2008), pp. 4777–4807.

[25] M. Stojnic, *Various thresholds for* ℓ_1*-optimization in compressed sensing*, arXiv preprint arXiv:0907.3666, (2009).

[26] P. Van den Berg and R. Kleinman, *A total variation enhanced modified gradient algorithm for profile reconstruction*, Inverse Problems, 11 (1999), p. L5.

[27] M. Wang, W. Xu, and A. Tang, *On the performance of sparse recovery via* ℓ_p*-minimization (*$0 < p < 1$*)*, IEEE Transactions on Information Theory, 57 (2011), pp. 7255–7278.

[28] W. Xu and B. Hassibi, *Precise stability phase transitions for* ℓ_1 *minimization: a unified geometric framework*, IEEE Trans. Inform. Theory, 57 (2011), pp. 6894–6919.

[29] W. Xu, M. Wang, J. Cai, and A. Tang, *Sparse recovery from nonlinear measurements with applications in bad data detection for power networks*, arXiv preprint arXiv:1112.6234, (2011).

[30] Y. Zhang, *A simple proof for recoverability of* ℓ_1*-minimization: go over or under*, Rice CAAM Technical Report *http://www.caam.rice.edu/∼yzhang/reports/tr0509.pdf*, (2005).